

# Loss Function: Binary Cross Entropy

Ki Hyun Kim

[nlp.with.deep.learning@gmail.com](mailto:nlp.with.deep.learning@gmail.com)

## Again, our objective:

- 세상에 존재하는 알 수 없는 함수  $f^*$ 를 근사하는 함수  $f_\theta$ 를 찾고 싶다.
- 따라서 함수  $f_\theta$ 의 동작을 정의하는 파라미터  $\theta$ 를 잘 조절해야 한다.
- 손실함수  $\mathcal{L}(\theta)$ 는 파라미터  $\theta$ 에 따른 함수  $f_\theta$ 의 동작의 오류의 크기를 반환한다.
- 따라서 손실함수를 최소화 하는 파라미터  $\theta$ 를 찾으면 된다.

# Binary Classification?

- e.g. 이 사람은 COVID-19 바이러스에 걸렸는가?
  - 입력: 키, 몸무게, 혈압, 혈중산소농도, 염증 수치
  - 출력: True / False
- 확률 문제로 바꿔 생각해보자.
  - remake: 이 사람이 COVID-19 바이러스에 걸렸을 확률은?

$$0 \leq P(y = \text{True}|x) \leq 1$$

$$P(y = \text{True}|x) = 1 - P(y = \text{False}|x)$$

# Because we use Sigmoid,

- Sigmoid의 출력 값은 0에서 1
- 따라서 확률 값  $P(y|x)$  으로 생각해볼 수 있음
- 실제 정답이 1이라면, 모델은 확률 값이 최대한 커지도록 학습될 것

$$0 \leq P(y = \text{True}|x) \leq 1$$

$$P(y = \text{True}|x) = 1 - P(y = \text{False}|x)$$

# Binary Cross Entropy (BCE) Loss Function

- $N$ 개의 vector들이 주어졌을 때의 수식
- $y_i$  의 값에 따라 수식의 왼쪽 term과 오른쪽 term이 on/off

$$\text{BCE}(y_{1:N}, \hat{y}_{1:N}) = -\frac{1}{N} \sum_{i=1}^N y_i^{\top} \cdot \log \hat{y}_i + (1 - y_i)^{\top} \cdot \log(1 - \hat{y}_i)$$

# Summary

- Regression의 경우에는 보통 손실함수 `MSELoss`를 활용하여 파라미터를 최적화
- Classification의 경우에는 `BCELoss`를 활용하여 파라미터를 최적화하며, Accuracy를 통해 우리는 모델의 성능을 평가할 수 있다.
  - `BCELoss`의 경우에는 확률/통계, 정보 이론과 밀접한 관련이 있음